White Paper Report

Grant number: HD-248511-16

Title of Project:
Project Andvari: A Digital Portal to the Visual World of Early Medieval Northern Europe

Principal Investigators:
Worthy Martin (University of Virginia)

Project Co-Directors:
Lilla Kopár (The Catholic University of America)
Nancy L. Wicker (The University of Mississippi)

Project Assistant:
Joseph Koivisto (University of Maryland)

Grantee Institution:
University of Virginia

Date of Submission:
February 15, 2019

Project Activities

The primary objective of the NEH Digital Humanities Level II Start-Up Grant was to develop a digital portal to provide integrated access to dispersed collections of Northern European art and artifacts of the early medieval period (4th–12th centuries). The main project focus was established through our Stage I grant period (July 2013–December 2014), funded by the NEH ODH (grant no. HD-51640-13). Guided by this overarching project goal, the specific **development objectives** have been as follows:

1. To address the challenge of providing access to the existing but dispersed artifactual record of the identified period and geographic scope;
2. To augment and enhance existing data and metadata;
3. To allow humanities scholars to study aggregated materials in an interdisciplinary fashion, thereby fostering innovative insights and analyses;
4. To promptly disseminate information about new discoveries, interpretations, and resources, and thus provide a shared virtual workspace for researchers and the public.

These objectives were established as a guiding framework for the actual planned activities of the project, namely the development of a free digital platform that will serve as an access point to the visual cultural heritage of the northern periphery of early medieval Europe as expressed in metal, stone, textiles, and other media by aggregating existing metadata records and digital surrogates of objects maintained within collections of third-party institutions.

Description of major activities

Funding for Stage II of Project Andvari was awarded to begin in August 2016, based on a proposal submitted by the University of Virginia. The initial Stage I grant was awarded to joint principal investigators, Kopár (The Catholic University of America) and Wicker (The University of Mississippi), and was administered by The Catholic University. As the Stage II grant shifted focus to active development of the *Andvari* platform, the principal investigating institution was changed to the University of Virginia in order to take full advantage of software development personnel at the Institute for Advanced Technologies in the Humanities (IATH). The role of PI was taken on by Worthy Martin, long-time project collaborator, technical expert, and director of IATH. Technical support and programming has been provide by senior programmer Robbie Bingler and other IATH staff members.

1. Metadata model

Initial development began with creation of a functioning **metadata model** to be used in the normalization of incoming records from third-party institutions. For Stage II, this was limited to a small set of partnering institutions: the British Museum, the Swedish National Heritage Board, and the Norwich Castle Museum. Based on the work of our Stage I grant, the metadata model was designed based on data elements that

were determined to be of most use to researchers when searching for artifacts through digitally-mediated platforms. This objective proved challenging as the task required the team to consider approaches to combining heterogeneous data sets from disparate institutions into simplified data models that would allow for consistent metadata display and comparison of disparate physical objects. Based on input from researchers, scholars, and, cultural heritage specialists, the Andvari team developed this initial model as a means of satisfying the greatest number of researcher needs while avoiding an overabundance of confusing data that would inhibit the efficient use of the platform.

2. Harvesting protocols and data harvesting

Concurrent with the development of the metadata model, work began on the development of **metadata harvesting protocols** for our partner institutions. As each institution maintained different metadata formats and data publishing practices, this required a considerable effort in producing redundant methodologies to serve each institution's data sets. These activities are described below:

- **British Museum (BM):**

  At the outset of Stage II, the BM hosted an easy-to-use endpoint that supported public reuse of metadata records through structured queries that returned object identifiers and related controlled terminologies that made data identification and harvest extremely simple. Using this platform, the team was able to develop a query to select resource identifiers of objects that relate to the project's chronological and geographic scope. With this list, we were then able to produce a custom script to harvest museum records from the public-facing collection's pages that would facilitate easy ingest into our custom-developed data model.

  Unexpected challenges, resulting changes and solutions

  While the initial metadata harvesting methodology proved effective, internal changes to the BM team responsible for administering and maintaining the public-facing data endpoint presented the team with numerous administrative and technical challenges.

  a) The data endpoint substantially altered the BM platform, causing widespread confusion about the persistence of uniform resource identifiers (URI) from the BM's datasets and concern about the harvestability of BM data for Andvari via automated methodologies.

  b) The instability of the BM data has been further complicated by the uncertain nature of their new platform, with frequent unscheduled outages impeding access and data linking.

  Fortunately, our initial harvest provided an adequate data set for our initial pilot. Despite concerns on the future usability and openness of BM data and linked open data URIs, we determined that a snapshot of BM data provides a worthwhile data set, especially when considered in the context of additional third-party data that would be present in the *Andvari* pilot.

- **Swedish National Heritage Board (SNHB):**

  Via the K-Samsök endpoint (also referred to as Swedish Open Cultural Heritage [SOCH]) (http://www.ksamsok.se/), we were able gain access that allows for querying and retrieval of cultural heritage data. Data is returned per user specifications for search specifications and data formatting. We were able to retrieve data for cultural artifacts meeting the following requirements:

  - Created between 300 and 1200 CE
  - Created during the following named periods:
    - Viking Age
    - Vendel Period
    - Migration Period
    - Early Middle Ages
    - Early Iron Age
    - Roman Iron Age

  This harvest process proved successful from the beginning of our efforts and is easily reproducible for future iterative harvests.

- **Norwich Castle Museum (NCM):**

  The Norwich Castle Museum, unlike the BM and the SNHB, does not host a public-facing data portal through which to harvest metadata records. Through partnership with representatives of the NCM, the project team was able to receive metadata and image files for museum resources, allowing for easy integration with our project data model.

3. Custom crosswalks and initial application infrastructure

With receipt of source data from our pilot organizations, the Andvari project team developed **custom crosswalks** for each of the data sources that support the structured extraction, transformation, and load of third-party data into the *Andvari* data model. Following the development of metadata harvesting techniques, developers at IATH began development of the **application infrastructure** that would serve as the foundation for the *Andvari* platform.

4. Review of data ingest and corrections

Data ingested was integrated into our data layer and reviewed by project team members for successful ingest and overall relevancy to the project. This **review** allowed us to identify procedural issues related to metadata mapping from source materials to our project-developed data model.

During initial review, several objects were found in our *Andvari* data set that fell outside of the defined chronological and geographic scope of our project focus. For example, project members found numerous records and images for nautical models (ships) produced in the mid-to late-20th century.

Following this initial assessment, we revisited our metadata extract scripts to ensure that we were not inadvertently selecting resources based on inaccurate data. After a **refreshed data extract**, we considerably reduced the occurrence of errant records.

5. Public interface design

After assessment of the preliminary database, the project team began designing the interface for the publicly-accessible *Andvari* web platform. Basing our design on existing collection interfaces that serve similar collections of cultural heritage resources and metadata, Koivisto, Kopár, and Wicker developed initial **design mock-ups** to serve as general guides for the developer group at IATH.

Although the project team was overall satisfied with design ideas, we felt that the team's collective skills were insufficient to adequately implement a design that is both cohesive and functional. To that end, the project team used remaining project funds to procure the assistance of a professional design consultant with a history of consulting on IATH initiatives. Using the project team's mock-ups as a starting point, the designer developed a comprehensive design scheme that established a unified graphic vocabulary for all pages of the *Andvari* platform. Attention to detail in developing this design scheme helps to ensure ease of use across all platform functionality, ensures consistent navigation through identifiable platform interactions, and promotes platform recognition for users.

6. Revision and restructuring of iconographic thesaurus

Continuing on the work of the initial grant period, the project team enhanced the **iconographic thesaurus** that was initially generated as a product of our Stage I grant. Reviewing our initial work and further assessment done with the help of a graduate student, the project team revisited the initial structure of the thesaurus, restructured terminologies and hierarchical associations as needed, and created a **second draft** of our proposed thesaurus. The enhancements to the thesaurus ensured a more accurate resource focusing on pre-Christian iconography in early medieval Northern Europe.

Following enhancements to the thesaurus, an **initial concordance assessment** was completed on data sets harvested from pilot institutions, aligning existing metadata describing the iconographic content of resources with the project-generated thesaurus in order to support easier collocation of resources via iconographic description and the pilot platform's iconographic search function.

7. Pilot platform prototype

The **pilot platform**—a confluence of our project data model, harvested data, iconographic thesaurus , underlying platform infrastructure, and specialized design schema—was ultimately realized in prototype format, hosted on UVA servers, and publicly available at http://hero.village.virginia.edu/andvari/. This platform serves as a home through which collections from our pilot institutions' collections can be easily discovered and collocated. Users first encounter the *Andvari* homepage, a beautifully designed landing page that greets newcomers with an image carousel, a tally of total objects on the platform—currently over 816,000—and useful guides on project overview, platform usage, and participation under development.

Search navigation is supported through three unique platform functionalities:

- **Keyword search:** This basic keyword search supports regular search practices by allowing users to search for select terms across all available metadata fields. This functionality will align with most users' expected platform behavior by providing a "single search box."
- **Advanced search:** The advanced search functionality provides more seasoned researchers and motivated users with a means of searching collections along more advanced logical means according to Boolean operators. Users can enter search terms for the following metadata fields:
    - Name
    - Location
    - Institution
    - Material
    - Period
    - Culture

  Each metadata element can support multiple search terms through the use of the plus sign (+) button next to each field, defaulting to a Boolean OR statement combining them.
- **Iconography search:** The iconography search functionality supports semantic searching based on applied iconographic description of objects per the *Andvari* iconographic thesaurus. Users can select headwords that then display the terminological hierarchies that exist beneath them. Users can select multiple terms, opting for either Boolean OR or AND combinations.

Publicizing *Project Andvari*

1. Presentations

Several presentations are promoting awareness of Project Andvari. Joseph Koivisto and Lilla Kopár gave a presentation about the project in the Lightning Round Presentations at the *NEH Digital Humanities Grant Directors' Meeting* in Washington, D.C., on September 16, 2016. Nancy L. Wicker spoke on "Andvari: Bridging the Gap across Disciplines and Institutions," in the Medieval Studies Program lecture series titled "The Virtual Middle Ages: A User's Guide" at The University of Mississippi, on February 7, 2019. Lilla Kopár will give a presentation about Project Andvari at the meeting of the International Society of Anglo-Saxonists (ISAS) in Albuquerque, New Mexico, on August 2, 2019. Kopár has also submitted a paper proposal (pending acceptance) to present the latest developments of the project at The Catholic University of America Research Day in Washington, D.C., on April 9, 2019. These presentations have given (and continue to give) project staff opportunities to answer direct questions about the project and the proposed platform, thereby promoting engagement within the disciplinary field while also increasing public anticipation of the coming product.

2. Publication

Joseph Koivisto, Lilla Kopár, and Nancy L. Wicker wrote a co-authored chapter "Bridging the Gap: Managing a Digital Medieval Initiative Across Disciplines and Institutions," published in *Meeting the*

Medieval in the Digital World, edited by Matthew Evan Davis, Tamsyn Mahoney-Steel, and Ece Turnator (Leeds: ARC Humanities Press, 2018, 223–240). This publication is helping to further expand project awareness and share our experience in various aspects of digital humanities with the wider world.

3. Crowdsourcing and social media

In addition to the primary project website (www.andvari.org), the Andvari team has employed several approaches to publicize the project and the proposed platform, garnering considerable interest from several researchers and institutions in the US and abroad.

While waiting to hear about the outcome the application for this NEH-ODH Level II grant, the Andvari team participated in a MicroPasts crowdsourcing project of The British Museum and the Department of Archaeology at University College London. The goal was to test the usability of the iconographic thesaurus being developed for the AndvariProject by tagging a sample set of 250 objects using a simple interface by which users could view a sample object from either the British Museum or Kringla (SNHB) and select terms from the supplied RDF thesaurus that appropriately describe the iconographic content of an object. Data collection and evaluation of the MicroPasts project completed in 2016 assisted our team in gathering data to assess the draft thesaurus. Simultaneously, the Andvari Project benefited from the wide exposure on the MicroPasts site. http://crowdsourced.micropasts.org/app/andvari/.

The project blog (https://projectandvari.wordpress.com) has generated over 2,400 views during the project period and has provided an excellent platform for disseminating project updates. The use of Twitter—through the @ProjectAndvari account—has also served as a project outlet, allowing for quick updates to over 300 individual followers. The project blog has received interest from several countries, including the United States, the United Kingdom, Sweden, Norway, Germany, Russia, and Belgium. Furthermore, a project logo was designed by Hungarian graphic designer Péter Sávai continues to provide visual branding for our project.

Accomplishments

The Andvari team accomplished the primary objective of the grant application to develop a workable pilot platform of a digital portal to provide integrated access to dispersed collections of northern European art and artifacts of the early medieval period (4th–12th centuries). A considerable amount of administrative effort and platform development was completed, and despite substantial setbacks during the September 2017–January 2018 phase of the grant period, we met all stated project goals. The team completed design of the data model, harvested data sets from three sources, ingested data into our PostgreSQL data infrastructure, and developed an interactive user design for the web portal.

Proposed Objectives from Initial Application

In order to accurately reflect the accomplishments of Project Andvari, the following list directly addresses the proposed objectives of the project period and the specific accomplishments to which they pertain. The stated objectives include:

1. To address the challenge of providing access to the existing but dispersed artifactual record of the identified period and geographic scope:

During the grant award period, the Andvari team developed the pilot digital platform that serves as an access point to the visual cultural heritage of the northern periphery of early medieval Europe by aggregating existing metadata records and digital surrogates of objects maintained within collections of three third-party institutions: The British Museum, the Swedish National Heritage Board, and Norwich Castle Museum. Access to the dispersed information from these three institutions is now freely available to users. This is the jumping-off point to include metadata records from more collections in the future.

2. To augment and enhance existing data and metadata:

To augment the existing data and metadata that were ingested into our digital platform, the team developed a standardized, controlled vocabulary to describe pre-Christian art of the northern periphery of Europe. This vocabulary is a needed addition to the previously described, largely Christian terminology of the Linked Open Data of the Art and Architecture Thesaurus of the Getty Foundation (http://www.getty.edu/research/tools/vocabularies/aat/) and also the coverage of the Princeton University's Index of Medieval Art (https://ima.princeton.edu/).

3. To allow humanities scholars to study aggregated materials in an interdisciplinary fashion, thereby fostering innovative insights and analyses:

From the outset of the project, the Andvari team was constructed to promote interdisciplinary engagement with the task at hand. Ongoing cooperation with information technologists, various scholars of the medieval period, museum professionals, and librarians has helped to ensure that project planning was considered through a diverse set of disciplinary perspectives.

4. To promptly disseminate information about new discoveries, interpretations, and resources, and thus provide a shared virtual workspace for researchers and the public.

The pilot platform will continue planned regular data ingest from contributing institutions, for example, from the Portable Antiquities Scheme, to which new discoveries of artifacts from England and Wales are constantly added. As a guiding principle, disseminating new information will help to inform platform development activities such as identification of important metadata elements, methodology and frequency of automated data harvesting, and generation of innovative discovery techniques to help users locate the most up-to-date object records.

Audiences

The initial users and primary audience of the proposed digital tool will be scholars and advanced students of various academic disciplines (archaeology, art, history, social sciences, literature, religion, etc.), as well as cultural heritage and museum experts. In addition, database owners who are eager to share their collections will be able to contribute to the project. As the project matures, it will be available for pedagogical use and interested members of the public. The scope is international and the project is intended to be broadly accessible. It will harvest resources and records from several institutions within the US and Europe, and eventually information will be made available in several languages, beginning with English, German, and Swedish, with more languages to follow.

Andvari seeks to introduce new ways to access information through aggregation and alignment of existing data, promoting research in ways that have previously remained unobtainable. For the scholarly audience, the result will surpass the existing access to varied resources through individual holding institutions. Although the project has yet to draw a true audience as the development of the pilot platform has just now been accomplished, it has garnered interest and participation from a wide variety of scholars in different parts of the world through the project blog, the Micropasts project, and international presentations about the project by team members during the NEH-ODH Level I and Level I granting periods. At its core, Project Andvari is a collaborative effort and has greatly benefited from the engagement of numerous representatives from a wide variety of institutions.

During initial research into extant platforms and authority resources, Project Andvari identified an ongoing project whose stated objectives aligned with our own in such a way as to promote collaborative efforts. The NEH-ODH sponsored PeriodO (http://perio.do/) project has expressed interest in collaborating with Project Andvari to enhance our ability to link and search across historical periods in our object records. Archaeological and art historical material is often categorized by period, but the dates of those periods vary from place to place, thus rendering chronological association and collocation difficult or, in many cases, impossible.

Evaluation

Preliminary evaluations of metadata harvesting protocols has produced mixed results on the efficacy of our approach. First and foremost, disruptions in the availability and reliability of data from the British Museum cast serious doubt on the stability of our proposed platform as one of our major data contributors could not be trusted as a reliable source of structured, publicly-available data. With this in mind, it has become apparent that *Andvari*, as well as any digital humanities initiative reliant on datasets from third-party organizations, would do well to formalize data sharing agreements via a memorandum of

understanding rather than trust in the reliability of publicly available data endpoints. This issue created considerable backlog for the project. Furthermore, variations in descriptive practices between pilot institutions and limitations on API search and retrieval protocols produced many inaccurate data results for our initial harvesting activities. While the team did their best to limit the possibility of introducing incorrect resource records to the initial *Andvari* dataset, we were unable to avoid all errant records as the initial record count—over 800,000 items—is too numerous to individually check. Future development on *Andvari* should include further assessment of the sustainability of our data harvesting and quality assurance approaches.

As a collaborative DH project across academic fields and various types of institutions, Project Andvari has proven to be an extremely valuable experience to all collaborators. Although crossing the disciplinary and institutional divides had sometimes presented us with challenges, the project provided exposure and training for the team members across disciplines, established institutional and personal connections that benefit the contributors beyond this very project, and (in the earlier phases) provided a number of graduate students with hands-on project experience in the digital humanities.

Continuation of the Project

In May 2019, we plan to apply for a NEH Digital Humanities Advancement Grant since we are ready for full implementation of *Project Andvari*. We plan to continue testing the functioning pilot using the collections of three international partner institutions (The British Museum, the National Heritage Board of Sweden, and Norwich Castle Museum) and also to extend the project by making agreements with further institutions and collections. We will focus on implementation of the platform with additional partners as well as sustainability of the overall *Project Andvari*. We also hope to work with the Institute of Museum and Library Services (IMLS) to advance access to, use of, and engagement with digital collections and services of museums. If our application to NEH is not successful, we are determined to apply for other sources of funding.

Longer Term Impact

Project Andvari has created the pilot for a free digital platform that currently serves as an access point to the visual cultural heritage of the northern periphery of early medieval Europe by aggregating existing metadata records and digital surrogates of objects maintained within three collections of third-party institutions. At the next stage, the project will extend that offering to the collections of more institutions. Ultimately, the project will facilitate interdisciplinary research in art, archaeology, history, and literary and religious studies, allowing users to study visual culture across media and beyond traditional

geographical and disciplinary boundaries. The innovative application of aggregated search methods and enhanced metadata will promote discovery and comparative analyses of artifacts in ways that have not previously been feasible. The new digital tool will offer aggregated search options of linked open data and metadata enhanced by public contribution through tiered access and will provide innovative engagement for both experts and the general public through access-controlled data enhancement features.

The successful launch of our pilot, and later the full-scale digital tool itself, will further illustrate the importance of such a resource to the growing field of digitally-enabled medieval research by showing that such a platform is technically feasible and capable of generating meaningful enhancements of existing data, and can inspire new research questions and cross-disciplinary collaborative projects.